REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
08-09-2015	MS Thesis	-

4. TITLE AND SUBTITLE
Models of Conflict Expression

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W911NF-12-C-0002

5c. PROGRAM ELEMENT NUMBER
1D10BP

6. AUTHORS
Paulo F. Gomes

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Raytheon BBN Technologies Corp.
10 Moulton Street

Cambridge, MA 02138 -1119

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
60679-NS-DRP.7

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Conflict is a critical element of social interaction. More and more it is being mediated by technology such as email, SMS and online texting. Consequently, it is becoming increasingly important to develop algorithms that could be applied to this growing data form. This report describes a predictive model for conflict strategy choice given demographic and personality input. It used crowd-sourced data concerning everyday life scenarios. A naive bayes classifer and support vector machine both perform above random. Additionally, statistical analysis of the data was consistent with previous research

15. SUBJECT TERMS
social interaction, conflict strategy, machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			William Ferguson
UU	UU	UU	UU		19b. TELEPHONE NUMBER 617-873-2208

Report Title

Models of Conflict Expression

ABSTRACT

Conflict is a critical element of social interaction. More and more it is being mediated by technology such as email, SMS and online texting. Consequently, it is becoming increasingly important to develop algorithms that could be applied to this growing data form. This report describes a predictive model for conflict strategy choice given demographic and personality input. It used crowd-sourced data concerning everyday life scenarios. A naive bayes classifer and support vector machine both perform above random. Additionally, statistical analysis of the data was consistent with previous research.

Models of Conflict Expression

Paulo F. Gomes

Reading Committee
Arnav Jhala
Michael Mateas

December 3, 2014

Abstract

Conflict is a critical element of social interaction. More and more it is being mediated by technology such as email, SMS and online texting. Consequently, it is becoming increasingly important to develop algorithms that could be applied to this growing data form. This report describes a predictive model for conflict strategy choice given demographic and personality input. It used crowd-sourced data concerning everyday life scenarios. A naive bayes classifier and support vector machine both perform above random. Additionally, statistical analysis of the data was consistent with previous research.

Contents

Chapter 1

Introduction

> We do not have the option of staying out of conflict unless we stay out of relationships, families, work and community. [Hocker and Wilmot, 2013]

We fight with our peers over cubicle spaces, with our relatives over what career path to take, with our partners on furniture arrangement. We participate in conflict from kinder garden quibbling over toys, to old age bickering over care taker competence. With its critical importance in human interaction, it is no surprise that it is the cornerstone of classic storytelling. Aristotelian drama thrives on conflict and we can see its influence on modern pop culture ranging from television to video games.

In the context of the current work, we consider interpersonal conflict to mean:

> A situation in which two individuals have opposing interests and at least one of them acknowledges said interests.

As more of our social interaction is mediated by technology such as email, SMS and online texting, so is interpersonal conflict. Social network companies have access to a large quantity of this type of data, even if its use is limited by terms of service. There is a trend of western mid-upper class teens and adults relinquishing their personal information by using social network apps. For instance, users of the latest Android facebook application have granted access to all their text messaging, a permission introduced early 2014. According to Facebook, in June 2014 there were 654 million mobile daily active users on average [Facebook, 2014]. Thus it is increasingly relevant to develop systems that can automatically analyze data. Given the importance of interpersonal conflict in social interaction, I believe that it should be an important aspect of such as system.

Although there has been extensive study of interpersonal conflict in the context of enterprise management and marital counseling, there is less research on modeling interpersonal conflict computationally. The potential applications are uncountable. For example, a virtual agent could detect that a conflict is arising between two friends and try to provide advice, present descriptions of similar recorded conflicts that match the situation, help the parties explore the results and consequences of different conflict strategies, or suggest a book in which the main character faces a similar situation. Modeling conflict could allows us to create more engaging and accurate training simulations by taking into account the social state of other agents with competing goals. Non Player Characters in video games could act more believably if they took into account the conflict context with the player.

In this project I focused on conflict strategy choice, that is, how people address conflict they are faced with. More specifically, I develop a conflict strategy prediction system: given a conflict situation description, and a person description, make a prediction of what conflict strategy is

the person most likely to choose. The study used crowd-sourced data concerning everyday life scenarios [Swanson and Jhala, 2012]. The corpus has responses to hypothetical conflict situations, person descriptions (demographic and personality self-assessment), and labels for the the type of the conflict strategy used. The implicit hypothesis was that one or more of these features would determine conflict response. I was also interested in considering different conflict sources.

Chapter 2

Background

Since the focus of this project is on conflict resolution strategy choice, I present a categorization that has been widely used in related work [Kilmann and Thomas, 1975]:

- **Dominating**: the individual pursues personal goals with low concern for the interests of the other individual. For instance, consider that a manager is contacted by a subordinate that tells him that he probably won't be able to meet a defined deadline. A dominating strategy by the manager might be to threaten to fire him if he does not meet the deadline.
- **Integrating**: the individual tries to account for personal but also other party's interests. In the previously described scenario, the manager might suggest a bonus in compensation for over time, and at the same time try to discuss with the subordinate the underlying reasons for the problem.
- **Avoiding**: the individual does not address either of the goals. In the same scenario, the manager might tell the subordinate to come back later because he is currently too busy.
- **Accommodating**: the individual thwarts personal goals and addresses other party's goals. In our scenario, the manager could simply tell the subordinate that it is fine that the deadline is missed without looking into the reasons for the schedule slide.

A Compromise category is also presented in [Kilmann and Thomas, 1975] as intermediate compared to the other four. In an initial modeling effort I considered that it would be more valuable to only use the four strategy types with clearer bounds. One can see a representation of the different strategies in Figure 2.1. These conflict resolution strategies have been correlated with personality traits. Extroverts tend to be more integrative [Kilmann and Thomas, 1975] [Myers, 1962] [Costa and McCrae, 1985] and dominating, while using less the avoiding style [Costa and McCrae, 1985]. Other results are presented in [Costa and McCrae, 1985]: Conscientiousness was also positively correlated with the integrative style and negatively with avoiding; Neuroticism, on the other hand, was negatively correlated with dominating and positively with avoiding. Lastly, openness was correlated with the integrative style. All these results will prove relevant in our statistical analysis of the dataset.

Conflict responses have also been classified as *Passive/Active* and *Aggressive/Non Agressive* in [Swanson and Jhala, 2012]. Definitions for both binary labels are shown below:

- "**Active** expresses whether the individual's action is an active step or is directly acknowledging the conflict."

- "**Aggressive** specifies whether the response was hostile towards the other party, for example, shouting or intimidation."

In order to further contextualize the characteristics of the gathered data set and related work it is necessary to consider what sources of conflict there could be. Here are five categories that I would like to highlight:

- **Conflict of Values**: two parties disagree due to opposing values or ideologies (e.g. forum post authors discussing their inconsistent views on gay marriage).
- **Goal Conflict**: individuals have different desired outcomes of a situation (e.g. a teenager wants to pursue an art career but her father is pushing for engineering).
- **Conflict of Interest**: two parties want to allocate scarce resources in different ways (e.g. two managers want to be section supervisors).
- **Affective Conflict**: when cooperating to solve a problem the participants realize that their feelings are inconsistent (e.g. roommates try to discuss a cleaning schedule and one starts crying every time a suggestion is directed at him).
- **Institutionalized**: the individuals follow predefined rules of interaction (e.g. defense and accusation lawyers during a trial).

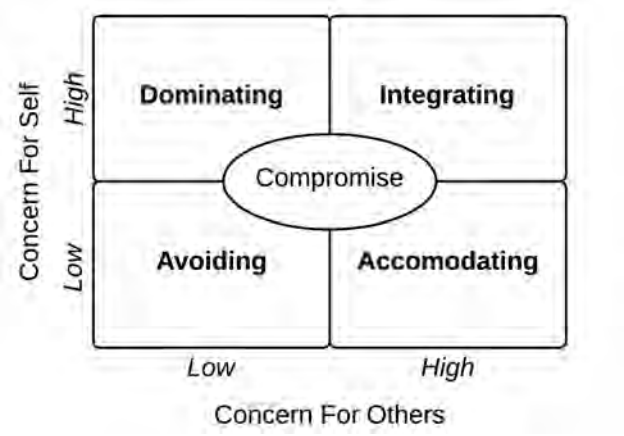


Figure 2.1: Conflict resolution strategy types.

Chapter 3

Related Work

While not specifically considering the concepts of conflict source or strategy type, stance recognition research has studied how internal constructs such as emotion and disagreement are expressed in textual form. Hence, applied to the right dataset, such techniques provide an implicit model of conflict. More explicit models of interaction can be found in frameworks that support intelligent virtual agents. When considering stories and scenarios with user interaction, conflict is an emerging concept. Finally, given the significant effort put in cleaning and organizing the used dataset, it is useful to look at what other publicly available datasets could be used for conflict modeling.

3.1 Social Interaction Modeling

Conflict has been modeled as a logical model of norms [Vasconcelos et al., 2009], as partial order causal link planning [Ware and Young, 2011], as part of an emotional appraisal process [Campos et al., 2013], as social games [McCoy et al., 2010] and using reactive planning [Mateas and Stern, 2004]. We will focus on approaches that consider external performative conflict since we are interested in the expression of conflict in text. I describe in more detail the FAtiMA [Dias et al., 2011] and ABL [Mateas and Stern, 2004] frameworks since I used them in my previous work concerning conflict.

CPOCL

In [Ware and Young, 2011] a system is presented that is able to generate conflict stories using partial order causal link planning. The actions taken by actors in the story constitute the steps of the plan, the partial order determines that certain steps must be performed before others, and the causal links indicate that one step made the precondition of another step true. A search is performed in the space of all possible plans with the refinement and maintenance of a partial plan. Actors' intentions are modeled as modal predicates. A conflict is a situation in which one actor has the intention of performing a step that will thwart the causal link relevant to the fulfillment of the intention of another actor. Conflicts are not necessarily pursued but rather can happen during the construction of the plans without invalidating them. The focus of this work is more on generating interesting stories rather than to model accurately how humans interact in real situations and make predictions.

Figure 3.1: FearNot! anti-bullying application screenshot.

FAtiMA

The computational model of appraisal theory of human emotions Fearnot AffecTIve Mind Architecture [Dias et al., 2011] has been used to model bullying scenarios [Aylett et al., 2007] that could be interpreted as a conflict situation. Figure 3.1 shows the anti-bullying application. FAtiMA considers that emotions are valanced interpretations of perceived events. Appraisal theories also claim that these interpretations depend on several appraisal variables: desirability of an event, praiseworthiness of an action, likability attitude towards an object, likelihood of a future event, among other. FAtiMA models this appraisal derivation process. Furthermore, it generates an emotional state based on these appraisal variables using an OCC theory of emotions [Ortony et al., 1990] inspired process (affect derivation). A strong motivation for having an agent architecture with a model of emotion is not only that it can simulate the emotional processes, but also, that emotion has been shown to be an integral process of human's decision making process [Damasio, 1994].

An important consideration in modeling interpersonal conflict in FAtiMA is that individual agents' behavior is defined by a STRIPS-like planner [Fikes and Nilsson, 1972]. Core to this planner are actions and goals. Actions have preconditions and effects. For instance, eating an apple might have as precondition that the agent believes it is holding an apple and as an effect that the apple is eaten. Goals have success conditions. For example, the agent Adam might have as a goal nourishment, with the success condition that it has recently eaten food. At any time point an agent has several potential goals to pursue in memory. The goal selected will depend on its relative importance. Importance is calculated according to the associated appraisal variables. For instance, if two goals are desirable, the one with prospective higher desirability will be selected. Authoring is performed editing seperate xml files.

FAtiMA architecture has actually been used to model the recognition by the virtual characters that there is a conflict [Campos et al., 2013]. However, there is little indication of given a specific scenario and character, what should the weights on the different goals and personality traits. Our approach tries to tackle this issue by learning this kind of weighting on real data.

Motivated by the important role emotion has in escalation (increasing the severity of conflict), the authors of [Campos et al., 2012] developed an IVA architecture that models conflict escalation in FAtiMA. There are three phases of conflict simulation: recognition, diagnosis and behavior selection. In recognition, actions and events that frustrate an agent's goal are identified and associated

Figure 3.2: Dream theater game screenshot.

with an urgency value; diagnosis, in which the cause (frustrated goal), participants, and relationships among them, are mapped to an emotion with certain intensity (increases with urgency). Finally, in behavior selection the agent chooses an action according to its personality strategy and emotional state. Personality strategy has two classes: attacking with high assertiveness, and evading with low assertiveness. Additionally, if the agent is in a negative emotional state, low cooperativeness strategies will be favored. The architecture was tested in a user study in which participants witnessed a video of IVA behavior modeled by the architecture in the My Dream Theater application (see Figure 3.2). Users reported identifying the agents' personality strategy and the simulated increase of conflict severity (escalation). In spite of these positive results, no strategy is presented regarding the architecture's parameter choice.

Cif

Comme il Faut [McCoy et al., 2010] is an AI architecture that tries to model social interactions through social games. These social games have several categories (e.g. Bully) that group similar situations. A social game has an initiator, a target, optionally a third party. Instances of the same social game will share success factors (is the initiator successful or not in its intent) and possible outcomes. These outcomes can be change in social status (e.g. two agents becoming friends). CIF's design is more focused on enabling scenarios that are compelling and provide a playable experience, rather than on psychological simulation.

One segment of the agents beliefs are the *social facts*. Social facts correspond to a history of what has happened. They are a vector of social game related events with a connected time. Another element of the AI's knowledge are the *social networks*. Also shared by all agents, each social network is a weighted directed graph that represents a certain aspect of the relationship between each two characters. Friendship, affection and respects are some of the network types that have been used in the past.

Finally social status rules are pre-conditions to a social status change that are defined as horn clauses. One might notice that there is no explicit encoding of personality. The corresponding personality is spread through the social facts and the cultural knowledge connections. The closest element are character traits, boolean attributes that can be used as wild cards in the architecture's processes.

At each agent's cycle, the following processes occur: goal setting, intent formation and social

game play. It attributes to each social game a volition (weight). The volitions are defined by summing the individual contribution of volition rules. Each rule contributes with its value if certain conditions are met. In the *intent formation* phase social games are ranked according to their volition. Given that the initiator and social game are defined, the *social game play* phase takes place. The outcomes of an individual social game are divided in positive (the initiator's intention is achieved), negative (the social game has no effects or counter productive effects) or neutral (the social game has no effects but it is likely that in future advances it will be successful). The social game will get a success value attributed to it based on a set of rules. These are similar to the ones mentioned in the goal setting, and when their preconditions are verified, each one contributes with its value to the total sum. Each of the categories has a non overlapping interval for its success values. If the game's success value belongs to that categories interval, the outcome will be of that category.

In spite of being able to model conflict with Cif, the weights set in the volition and social game rules need to be authored. The Prom is an example of a entertaining performance of the system [McCoy et al., 2010]. Nonetheless, the system does not give clues on how real people react in certain situations. The goal of Cif is much more an entertainment one than a behavior predictor one.

ABL

Another tool to model character expressive behavior is ABL, a programming language for reactive planning [Mateas and Stern, 2004] [Mateas, 2002]. It was designed to support virtual characters and it compiles to Java. It follows the principles of a hierarchical task network style planner. Goals are not defined by preconditions and success conditions. Instead, they are defined by the possible solutions for that goal. These solutions are called behaviors. Behaviors are recipes to reach the goal. A behavior itself has preconditions and children. These children can be atomic actions or new goals. Executing the behavior is performed by executing its children. At any time the system maintains a tree of the current active behaviors.

Like FAtiMA, ABL has atomic actions called acts. They can be declared actions in the world or mental acts. Mental acts are arbitrary pieces of Java code. Another crucial structure of the language is the Working Memory Element (WME). A WME is a piece of information representing a belief of the AI about the current state of the world or of a previous event. WMEs are the fundamental element of information representation in ABL. They can be referenced in preconditions. A behavior is only activated if there is a consistent unification of all the variables in the preconditions. Additionally, WMEs can be changed in mental acts.

In Façade, ABL was used to model the marital disputes between two characters Grace and Trip [Mateas and Stern, 2004] (see Figure 3.3). Additionally, ABL is currently being used to model social interactions in potentially conflicting situations [Shapiro et al., 2013]. In this case different perspectives of the situation are modeled by Social Games as in Cif [McCoy et al., 2010]. The social games define what are the main variables at stake (power, safety, friednliness, etc.). Additionally, rules define which is social game should be on the forefront for each character. The work is specially valuable in the dynamic modeling of social interactions. However, the weights given in the social rules have to be fine tuned and authored, and although it is possible for the authors to tailor them to a specific scenario, is hard to identify ways to define them for any scenario such as we propose.

Figure 3.3: Façade game screenshot.

Other

Multi-agent systems often consider the interaction between agents. For instance, COM-MTDP is a method to analyze the relation between optimality and algorithm complexity of agent teamwork [Pynadath and Tambe, 2002]. However, it assumes that there is a common goal which is definitely not the case in goal conflict as described in the background.

There has also been interest in conflict modeling for corporate tools support. PERSUADER is a system to support conflict resolution based on case base reasoning [Sycara, 1993]. It proposes strategies and helps the participants realize what are their main concerns. The system assumes that participants are fully engaged in the conflict. My task has more to do with classification and detection of conflict.

In [Sina et al., 2014] the authors explore the use of crowd sourcing in the context of serious games. They describe a system that is able to rewrite parts of a social interaction scenario description, sometimes involving conflict, by maintaining consistency with the rest of the scenario. This reworking uses crowd sourced everyday activity descriptions. The algorithm has three main steps: identifies which story elements need to be replaced with a Maximal Satisfiability Solver; uses k-nearest neighbor to match the scenario with an everyday example crowd sourced; finally it uses a natural generation system based on templates to regenarate the textual segments using the crowd sourced data. The system is more directed in filling in gaps with everyday descriptions rather than make predictions on how characters would react when confronted.

3.2 Stance Recognition

Beyond the just mentioned systems that emphasize generative power and explicit models of social interaction, it is important to consider discriminative models with a more implicit approach to the same themes. For instance, the objective of the work presented in [Misra and Walker, 2013] is to be able to identify a post response as being a disagreement or agreement independent of the forum topic. The study considers the Internet Argument Corpus [Abbott et al., 2011] manually annotated for disagreement/agreement with high inter-rater agreement. Conversations are abstracted as a sequence of speech acts (PROPOSAL, ASSERTION, ACCEPTANCE, REJECTION). Rejections are subdivided in several categories according to Horn's nomenclature [Horn, 1989]. In addition the authors propose two additional speech acts that do not entail direct logical inconsistency and result

in three new types of rejection: denying a communication of an assertion as a transfer of belief to a person, maintaining a belief is inconsistent with what is said, citing contradictory authorities. They created n-gram (bigrams and trigram were more informative) categories for agreement and denial (has more textual indicators) by studying a specific forum topic in the corpus. Also considered other features: cue words, duration of the post, hedges, polarity and punctuation. A decision tree was trained in one topic and tested in the remaining using all features. Compared with using non differentiated unigram and unigram+bigram approaches performance was significantly better. Doing a gain ratio analysis shows that there are discriminant n-gram disagreement independent features. In a ablation study punctuation and cue words were the most prominent and hedges were less than expected.

In [Hassan et al., 2010] the research goal was to identify if there is an attitude or not, and if it is positive or negative. In their method they start by only analyzing posts with second person pronouns, which might be filtering out relevant ones. They do a syntactic tree analysis of sentences only keeping the subtree starting at the second person pronoun. To find the polarity of words they generate a graph based on word net and some initial ground truth. Random walks are performed to extend polarity information to non-annotated words. Each sentence is mapped to 3 schemas: lexical, polarized words are replaced with NEG/POS tag, others stay the same; part-of-speech, words are replaced by part of speech tags; word sequence of the shortest path connecting second person pronoun to a polarized word. For each schema, and each mode (attitude, no attitude) a markov model is trained. Each node corresponds to a token, and the probability of transition between nodes corresponds to the ratio between the number of times it was witness by the number of times the starting node has been witness. The probability of specific sequence is then defined by the product of the probability of each individual transition, a process analogous to a bigram system. Then for each schema, they calculate the ratio between the log likelihood of the sequence being generated by the attitude mode as opposed to the non-attitude mode. A support vector machine is then trained using the three resulting ratios as features. For attitude polarity, the authors consider the average shortest path length between second person pronouns and positive words and compare it to the value relative to negative words. Both this and [Misra and Walker, 2013] article have potentially valuable feature suggestions for our prediction task.

Automatic detection of conflict has been studied in the context of complex human-machine interaction [Kanno et al., 2006]. In this work the model tries to encode false beliefs of team members when interacting with a machine. It relies on the assumption that there is a team intention resulting from the combination of individual member's intention. Additionally, it considers that conflicts result from the false beliefs. They describe a semantic logic for beliefs and intentions which emphasizes the potential importance of theory of mind in the conflict modeling context. They infer possible goals by: getting current valid goals in the interaction context; ground plans that can achieve them; match the members actions to all the plans that contain it; order the plans according to domain specific heuristics. Belief inference takes into account that other members might be working on a consistent plan or not. However, it appears to only be tractable in a highly standardized interaction. Furthermore, it is assumed that the task at hand is cooperative and that there is a limited number of easy to identify operands which might not be the case in goal conflict.

In [Cheong et al., 2011] authors define conditions in which a goal might arise and cause a conflict. They describe a method for detecting conflict through physical input. User conflict resolution choices contribute to a score on, assertiveness, cooperation and relationship. Conflict resolution aspects (trivial/important goal) contribute to a score of the 5 TKI resolution strategies. Authors separate the strategy initially chosen from actual resolution. This work differs from mine because identifying different strategy choices from the game world is more direct, than trying to detect such actions in everyday life scenarios.

3.3 Data Sets

Most of the previous section's methods require a considerable amount of data to be effective. In [Walker et al., 2012] the authors present the Internet Argument Corpus (IAC) comprising 390 704 forum posts in 11 800 discussions by 3 317 authors. It has potentially useful annotations in the context of conflict, disagreement and nastiness, having reasonably high inter-rater agreement. Furthermore, the research highlights discourse markers that were likely to appear in posts for disagreement and agreement. The disagreement markers being: *really* (67%), *no* (66%), *actually* (60%), *but* (58%), *so* (58%) and *you mean* (57%). The agreement markers were: *yes* (73%), *I know* (63%), *I believe* (62%), *I think* (61%) and *just* (57%). The dataset is distinct from mine because it mainly concerns opinion discussion between people that in many cases do not have to interact face to face. Sharing resources between participants in the forum is not as explored (conflict of interest). Nonetheless, it would be interesting to consider the discourse markers mentioned in our prediction task.

Controversy detection has been studied in the context of Wikipedia article edits, informally called Wiki Wars [Jankowski-Lorek et al., 2014]. The authors used trustworthiness annotations collected with the Article Feedback Tool to detect controversy. Trustworthiness was scored from 1 star to 5 stars. From 963 articles labeled as controversial by Wikipedia admins, 219 were selected having at least three evaluations, with 56% having more than 100 ratings. The final dataset also contained 219 non controversial articles. Applying a random forest algorithm with features extracted from these ratings resulted in a Area Under the Curve of 88%. Their dataset is available online at the datahub.io website [1]. The authors also present an emotion polarity based classifier for the wiki article talk sections. Here is a response to an article merge proposal for *Anarcho-capitlism*:

> `<li><b>Oppose</b>`
>
> for reasons mentioned above by editor Sharangi: ïn fact there are older and more popular market anarchist ideologies that are even anti-capitalist" but will support merger with either main Anarchism article or Anarchist Economics article. Note that this is not the first time someone has wanted to redirect this article, and that the last two attempts ended with abuse by ancap editors, which escalated to the noticedboards and several times required suspension from Wikipedia.
>
> `<a href=\"/wiki/User:Finx\" title=\"User:Finx\">Finx</a>`
> `(<a href=\"/wiki/User_talk:Finx\" title=\"User talk:Finx\">talk</a>)`
> `16:15, 4 September 2013 (UTC)</li>`

The article considers detection of interpersonal conflict caused by difference of opinion, and many cases ideology. Unfortunately, many of the editors do not share resources besides wikipedia page space (conflict of interest).

There are several social interaction data sets that do not have conflict related annotations. In the twenty newsgroups data set [Mitchell, 1997] there are a total of 20 000 messages with themes ranging from politics to hardware. Many posts have a single response, and the discussion concerns more opinion rather than the personal life of the participants. In [Galgani and Hoffmann, 2011] the authors describe a corpus of formalized reports of legal decisions in Australia. Although representing social interactions, the reports themselves are presented more as a narrated monologue, rather than a conversation. The emails between higher management in ENRON was open to public in the sequence of an investigation. A total of 2 200 emails were annotated as being business or personal.

[1] http://datahub.io/dataset/controversy-of-wikipedia-articles-using-aft

12% were consistently placed in the second category [Jabbari et al., 2006]. To be applied to our task, this subset would further need to be filtered to only include conflict situations and segmented for scenario onset and strategy choice. If after this process there is enough relevant data, it could be used to train a predictor.

Chapter 4

Previous Work

The present work follows from my interest in modeling expressive behavior of believable characters. I have developed a framework for agents with emotions and episodic memory [Gomes et al., 2011a], studied how improvements in believability can be accessed [Gomes et al., 2013], and compared the authoring process of different AI architectures [Grow et al., 2014]. Finally, regarding previous work more closely related to conflict, I have made an initial proposal on how to model conflict [Gomes and Jhala, 2013] and identified important features for a potential conflict strategy prediction system.

In that article I discussed how FAtiMA [Dias et al., 2011] and ABL [Mateas and Stern, 2004] could potentially be used to model conflict situations. For ABL we considered one ABL Entity encoding the whole conflict situation, a technique used in practice by the Games and Playable Media Group in the IMMERSE project [Shapiro et al., 2013]. We defined the following types of goals: *infer conflict* and *resolve conflict*. There was only one infer conflict goal that was used to detect all conflicts. There were different behaviors for this goal that represented different conflict situations (e.g. owing conflict). The behavior selected depended on the preconditions that encode the conflict context (e.g. a character owes another money). I created goals which represented trying to resolve each type of conflict situation (e.g. resolveChoresConflict). Furthermore, each resolve conflict goal had different behaviors corresponding to different resolution strategies (dominating, avoiding, accommodating, and integrating).

Contrary to ABL, using the situation as an abstraction is not an option in FAtiMA since an individual emotional state is maintained per agent and it affects the decision making process. In FAtiMA we defined that each goal corresponds to a different resolution strategy. Instead of selecting the strategy through preconditions, the strategy is chosen according to the different importance the agent gives to the goals: goals with higher importance for an agent are more likely to be pursued than those with lower importance. This importance corresponds to a desirability that in turn affects the agent's emotional state. Moreover, since goals do not have recipes on how to execute them, the atomic actions had to have a strategy bias, meaning that if a strategy is chosen, certain actions are more likely to be selected for execution. This was achieved by setting the value of a character attribute if an action is chosen, and using that same attribute/value pair in a goal success condition.

We found it harder to directly map theoric conflict concepts of organizational management, such as resolution strategies, in FAtiMA. Neither goals or unitary actions in FAtiMA are a good fit for the notion of strategy. On one hand strategies should define acted behavior like unitary actions, on the other they should also encode reasons dependent on character traits, that in FAtiMA can only be specified at the goal level. We were forced to use boolean character attributes to establish this link, which are little more than flags. In ABL we could map a resolution strategy (e.g. dominating) in a

type of conflict (e.g. owes conflict) to a behavior fulfilling a specific goal (e.g. resolveOwesConflict).

FAtiMA's focus on emotions does not match well the conflict theory we considered. Nevertheless, conflict situations tend to generate emotional responses and FAtiMA generated an emotional state for the characters with little additional effort (e.g. fear emotions caused by threatened goals). Concerning model checking, the fact that XML authoring errors in FAtiMA are only detected at runtime makes debugging slower. ABL's compiler error checking allowed a faster iterative process.

In regards to variability, when in ABL two behaviors fulfill a goal, have their preconditions met, and have the same specificity, the system selects one random behavior (e.g. choice between dominating or integrating strategies if the character has high concern for self). Thus, variability is embedded on how the behavior choice is made. In opposition, by default FAtiMA ranks plans according to the extent by which they achieve goals, selecting the optimal one. Consequently, at any time only one behavior can be selected, even if two should be equally likely. For instance, if goal A's importance is only slightly more important than B, there should be a close to 50% chance of B being selected, but currently A would have a 100% chance of being selected. There is still variability due to the dynamic influence of the emotional state on the characters decision making, and consequent numeric uncertainty, but it is harder to get an insight on how that variability will occur.

In a related topic, it is harder to fine tune the experience from a design point of view in a specific direction in FAtiMA because so much of the action choice is left to the emotionally driven planner (policy change). It is unclear how each specific numeric importance value, in the goals for instance, will affect the resulting actions (e.g. effect on behavior choice of changing the exact importance on goals). In ABL in by linking behaviors goals explicitly fine tuning is more flexible. Nevertheless, there is still some numeric obscureness when it comes to scalar values used in ABL's preconditions, since for an author to understand which behavior will be selected in a certain context she needs to go through all behavior preconditions fulfilling that goal. This emphasizes the importance of exploring data driven approaches to model social interaction.

Chapter 5

Conflict Strategy Prediction

Reiterating the objective of this project, I wanted to develop a conflict strategy prediction system: given a conflict situation description, and a person description, make a prediction of what conflict strategy is the person most likely to choose. For that purpose I used crowd-source collected data gathered in [Swanson and Jhala, 2012]. The corpus has responses to hypothetical conflict situations, person descriptions (demographic and personality self-assessment), and labels for the the type of the conflict strategy used. We will call the participants answering what they would do in a scenario responders. In summary, the construction of the original corpus consisted of the following steps:

1. collect narratives: crowd sourced workers were prompted online for a short textual description of an experienced conflict scenario including its outcome.

2. hypothetical scenario creation: outcome and personal details were removed from the scenarios. The authors changed the perspective from first person to second (e.g. *My friend wrecked my car ...* to *Your friend wrecked your car ...*).

3. collect responses: crowd sourced workers were prompted online for a short textual description of what they would do in a hypothetical conflict situation.

4. annotation: crowd sourced workers were prompted online to label collected responses regarding the type of conflict strategy chosen.

5.1 Data Description and Cleaning

Data concerning responders had a variety of features and can be divided in two categories: demographic and personality. Demographic data includes: sex, as male or female (no other tag was available); age, in 6 ranges (6-12, 13-18, 19-25, 26-40, 41-65, 66 or older); education, in categories (Never graduated high school, between high school and graduate[1], Graduated college, A graduate degree other than an M.D. or Ph.D., An M.D. or Ph.D.); number of cell phone numbers in ranges (No cell phone, 1-9, 10-39, 40-99, 100-219, 220-259, 260 or more); number range of text messages received per week (None, 1-9, 10-39, 40-99, 100-219, 220 or more); number of social network friends in ranges (None, 1-9, 10-39, 40-99, 100-219, 220 or more); number of physical exercise hours per week (None, 1-2, 3-5, 6-8, 9 or more); number video game hours played per day (None, 1, 2, 3, 4, 5

[1]The careful reader will notice that this category is too broad. In the original data two education values had ambiguous labeling. Unable to recover the specific labels I was forced to map the instances to a broad common category.

or more); number of tv hours watched per day (None, 1, 2, 3, 4, 5 or more); and textual description of closest town.

Personality was accessed using a short version of the Big Five Inventory [Rammstedt and John, 2007]. Participants had to answer 11 five point likert scale questions regarding their agreement with a statement of self-description (e.g. I see myself as someone who has few artistic interests). Each question answer is supposed to positively or negatively correlate with one dimension of Big Five (openness, conscientiousness, extraversion, agreeableness, and neuroticism) according to the instrument authors. I calculated a 1 to 5 (low to high) value for each dimension using the following procedure: for positive correlations I considered the actual value in the likert scale, and for negative ones 6 minus the actual value; for each dimension I averaged these scores obtaining a single value.

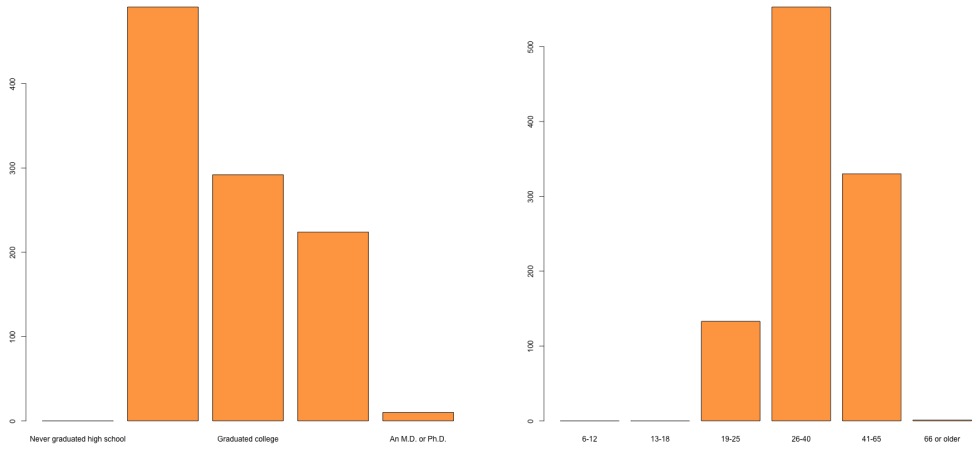
According to the data collection method, the core scenario attributes of original scenarios should have been maintained in the hypothetical ones [Swanson and Jhala, 2012]. When the original scenarios were gathered crowd workers were asked to input the type of relationship between the people in conflict as a categorical field: Stranger, Acquaintance, Romantically Interested, Friend, Romantically Involved, Close Friend, Spouse. As there are implicit relations between the labels, I created two additional numeric fields: a numeric relationship and a numeric involved. With numeric relationship I tried to encode the importance of the relationship with the following mapping: 1 - stranger; 2 - acquaintance; 3 - romantically interested OR friend; 4 - romantically involved OR close friend; 5 - spouse. With involved I tried to encode if the two main elements in the relationship were potentially involved with the following mapping: 2 - romantically involved OR romantically involved OR spouse; 1 otherwise.

Moreover, for the analysis done so far I considered two of the annotation dimensions on the responses: *active* and *aggressive*. Responses were annotated as being as Passive or Active, and orthogonally as Aggressive or Non Aggressive as defined in the Background. There are two annotation data sets: *user study*, with more annotations on a smaller number of scenarios and responses; *corpus*, with more scenarios and responses but less annotations per response. There is overlap between the response instances labeled in both. Since I mostly used *corpus* on the classification task and analysis, I will use annotations to refer to the corpus data set unless stated otherwise. For each dimension in corpus and each response we typically have 7 annotations. We consider the label to be the majority vote between annotations.

We will group up variables by data type. Ordinal features, are those for which order between positions is known but not the relative differences between positions [Field and Hole, 2003, pp. 7–6]. Ranges and self-reported likert scale are good examples of ordinal data. Consequently, we consider the following features to be ordinal, and will refer to them simply as *ordinal features*: age, education, number of mobile numbers range, number of mobile numbers range, number of text messages, number of social network friends, number of physical hours, number of video game hours, extraversion, agreeableness, conscientiousness, neuroticism, openness, and relationship. The *nominal features* (unordered categories) will be sex and involved. Lastly, since scenario description and city are both free form text, we will refer to them as *text features*.

For the reported analysis I performed an inner join merge of demographic, personality, scenario, response and annotation data. Given that to do predictions we would need as rich data as possible I decided to discard instances for which we did not have one of the mentioned subsets of data. Additionally, for some subsets of data I had to merge several file batches with slightly different table schemas. The code and data is publicly available at https://bitbucket.org/pfontain/conflict-data-cleaning.

The description henceforth will refer to the merged data. There were 164 different scenarios, a total of 1017 responses (a ratio of 6.2 responses per scenario), and 90 responders. Regarding the responses responders: 37% male responses, all having at least higher school education (details in



(a) education level (b) age

Figure 5.1: Responders' demographics

Figure 5.1a), 99% having ages between 19 and 65 (details in Figure 5.1b), 83% were labeled as Active, and 16% were labeled as Aggressive.

5.2 Statistical Analysis

The implicit hypothesis was that one or more of the features (scenario, demographic or personality) would determine conflict response. Consequently, we should expect differences of feature distribution between different response groups (e.g. Active vs Passive). We can further take advantage of results presented in the Background relating conflict strategy and personality [Kilmann and Thomas, 1975] [Myers, 1962] [Costa and McCrae, 1985]. As described previously, in interpersonal conflict integrative behaviors try to address both parties concerns actively. Therefore, I believe integrative behaviors should typically be perceived as active. By definition, avoiding behaviors should be considered more passive. Thirdly, due to low concern for others, dominating behaviors are more likely than not to be considered aggressive. If we take these assumptions, we can map previous results relating personality and conflict choice to the active and aggressive annotations.

- Extroverts tend to be more integrative, dominating and use less avoiding. Consequently, *extroverts should have responses that are perceived as active and aggressive.*
- High conscientiousness individuals tend to be more integrative and use less avoiding. Consequently, *Conscientious individuals should have responses that are perceived as active.*
- Neurotics tend to be less integrative, less dominating and use more avoiding. Therefore, *Neurotic individuals should have responses that are perceived as passive and non aggressive.*

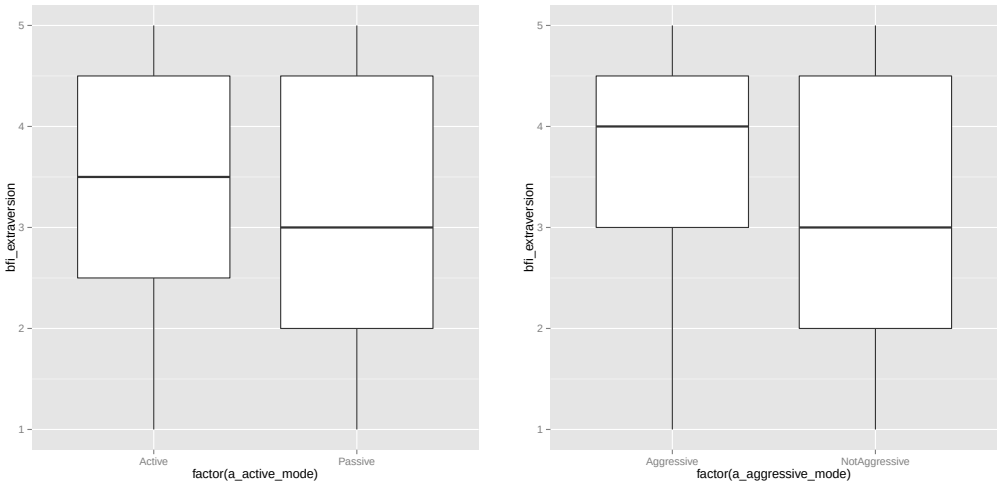
To verify these intuitions, I performed a Wilcoxon Mann-Whitney rank sum test for each ordinal feature, grouping responses by annotation label. I did this for active and aggressive labels. The test results (p-value and effect size) are presented in Table 5.1 and 5.2 together with medians for each feature/label. Features for which the effect size (r) is positive are those for which the ranking in Passive is lower than in Active, when the effect size is negative then ranking is higher for Passive.

	Mdn Act	Mdn Pas	p	r
age_group	4.00	4.00	0.0045	-0.0890
education	2.00	3.00	0.0003	-0.1145
n_mobile_numbers	4.00	3.00	0.0174	0.0746
n_text_messages	3.00	3.00	0.0031	0.0926
n_friends_sn	4.00	4.00	0.0078	0.0835
n_hours_pa	3.00	3.00	0.0002	0.1152
n_hours_vg_day	2.00	2.00	0.0566	0.0598
n_hours_tv_week	4.00	3.00	0.0767	0.0555
bfi_extraversion	3.50	3.00	0.0387	0.0648
bfi_agreeableness	4.00	3.33	0.0934	0.0526
bfi_conscientiousness	4.50	4.50	0.0127	0.0781
bfi_neuroticism	2.00	2.00	0.0158	-0.0757
bfi_openness	4.50	4.50	0.0954	0.0523
friendship_num	3.00	2.50	0.0305	0.0678

Table 5.1: Wilcoxon Mann-Whitney rank sum test for each ordinal feature, grouping responses by Active label on corpus data. labels: Mdn Act - feature median for active responses, Mdn Pas - feature median for passive responses, p - p-value double tailed, r - effect size.

	Mdn Agg	Mdn Not	p	r
age_group	4.00	4.00	0.6216	-0.0155
education	3.00	3.00	0.6692	-0.0134
n_mobile_numbers	4.00	3.00	0.1856	0.0415
n_text_messages	3.00	3.00	0.0949	0.0524
n_friends_sn	4.00	4.00	0.3941	0.0267
n_hours_pa	3.00	3.00	0.3251	0.0309
n_hours_vg_day	2.00	2.00	0.1782	0.0422
n_hours_tv_week	3.00	4.00	0.7039	-0.0119
bfi_extraversion	4.00	3.00	0.0000	0.1590
bfi_agreeableness	4.00	3.67	0.0588	0.0593
bfi_conscientiousness	4.50	4.50	0.0006	0.1072
bfi_neuroticism	1.50	2.00	0.0000	-0.1707
bfi_openness	4.50	4.50	0.0026	0.0944
friendship_num	3.00	3.00	0.4219	-0.0252

Table 5.2: Wilcoxon Mann-Whitney rank sum test for each ordinal feature, grouping responses by Aggressive label on corpus data. labels: Mdn Agg - feature median for aggressive responses, Mdn Pas - feature median for non aggressive responses, p - p-value double tailed, r - effect size.



(a) active (b) aggressive

Figure 5.2: Responses' responders extraversion grouped by labels.

There is a higher responder extraversion median on responses that were active ($3.5 > 3.0$) as can be seen in Figure 5.2a, as well as responses that were aggressive ($4.0 > 3.0$) as presented in Figure 5.2b. Regarding active annotation, we falsify the null hypothesis with $p = 0.019$ (one-tailed event). For the aggressive annotation, we falsify the null hypothesis with $p < 10^{-6}$. There is only a small effect size for active ($r < .10$) and a small to medium effect size for aggressive ($.10 < r < .30$). Considering effect size directives from [Field and Hole, 2003, p. 153].

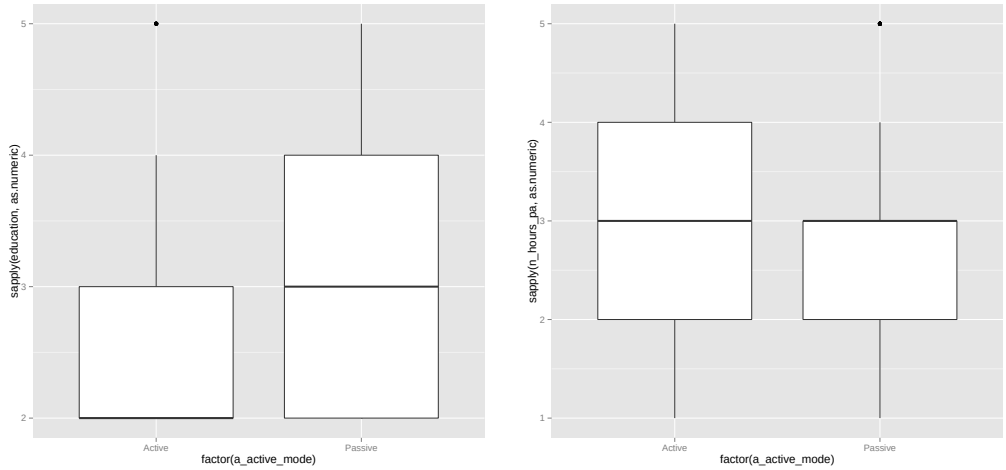
Besides extraversion, there is a significant difference ($p < 10^{-3}$) in the education level on responses that were active as can be seen in Figure 5.3a, with passive responses coming from higher education individuals with a small to medium effect size ($.10 < r < .30$). Passive responses also came from individuals with higher neuroticism ($p < 0.05$ and $r \sim .1$). In contrast active responses came from higher conscientiousness individuals with a significant difference ($p < 0.05$) and small effect size ($r \sim 0.08$), and having higher reported physical activity (Figure 5.3b) with a significant difference ($p < 10^{-3}$) and a small effect size ($r \sim .1$).

Grouping responses by aggressiveness, there is a significant difference ($p < 10^{-7}$) in the neuroticism level Figure 5.4a, with non aggressive responses coming from higher neuroticism individuals with a small to medium effect size ($.10 < r < .30$). Aggressive responses came from individuals with higher conscientiousness (Figure 5.4b) with a significant difference ($p < 10^{-3}$) and a small effect size ($r \sim .1$), and higher openness ($p < 0.01$ and $r \sim 0.1$).

Finally, I performed a correlation analysis between the response features previously highlighted. Table 5.3 shows the spearman correlations. I would point out the inverse correlation between neuroticism and extraversion, which indicates that in our sample we tended not to have neurotic extroverts.

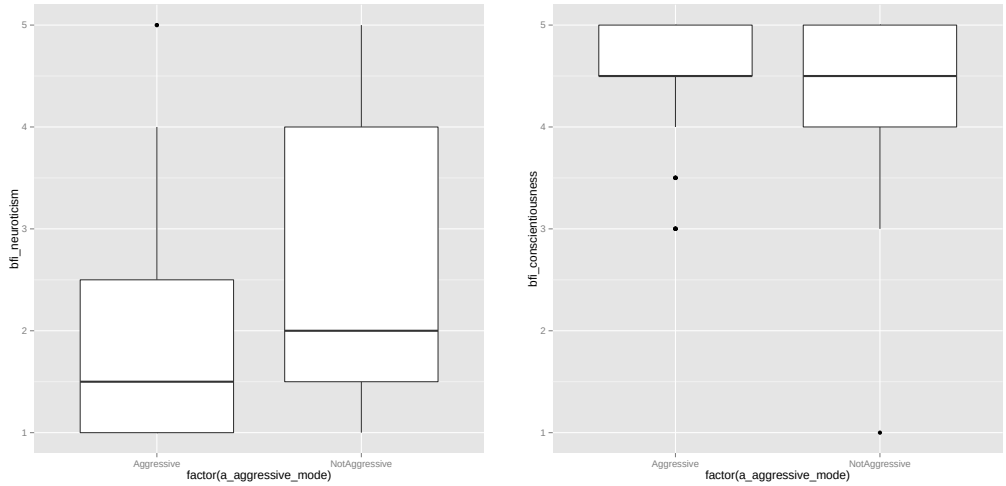
5.3 Prediction

Regarding prediction I will further focus on Active/Passive annotations. I split the corpus responses pseudo-randomly in train and test set (75%/25%) such that the ratio of active to passive responses was the same in both sets. I applied a naive bayes classifier and a Support Vector Machine (SVM) using the libSVM library [Chang and Lin, 2011]. The Naive Bayes classifier used ordinal and binary



(a) active (b) active

Figure 5.3: Responses' responders education and physical activity grouped by active label.



(a) aggressive (b) aggressive

Figure 5.4: Responses' responders neuroticism and conscientiousness grouped by aggressive label.

	educa	n_hou	bfi_e	bfi_c	bfi_n
educa	1.00	-0.41	0.0850	-0.1635	0.09
n_hou	-0.41	1.00	0.1809	0.2490	-0.26
bfi_e	0.08	0.18	1.0000	0.3197	-0.58
bfi_c	-0.16	0.25	0.3197	1.0000	-0.30
bfi_n	0.09	-0.26	-0.5841	-0.3029	1.00

Table 5.3: Correlations between features that were found to be statistically different across labeling groups. Labels: educa, education level; n_hours, physical exercise hours per week; bfi_e, extraversion; bfi_c, conscientiousness; bfi_n, neuroticism.

	NB	SVM
accuracy	**0.75**	0.59
recall	0.85	0.64
true negative rate	0.28	0.37
harmonic mean recall	0.42	**0.47**

Table 5.4: Performance metrics for Naive Bayes (NB) and

features, fitting gaussians to ordinal values and using frequencies for the binary ones. Cells with 0 probability were replaced by 0.001.

The SVM considered ordinal and binary features as numeric. I selected the radial basis function method (default) and used a class weighting that accounts for the class unbalance: $w_A = 0.2$, $w_P = 1$ ($0.2/1 \sim$ # passive responses/# active responses). In order to select gamma (γ) and cost (C), I performed a grid search with a 5-fold cross validation schema on the training set. The values obtained were the following: $\gamma = 2$ and $C = 2$. In Table

Since the classes are unbalanced ($\sim 10 : 2$), accuracy, recall and precision will be strongly weighted by the predominant responses (active). Thus looking at true negative rate is important ($tnr = tn/(tn + fp)$). Furthermore, if we average recall and true negative rate, we get a metric that weights both classes equally. Since we are averaging rates, I use the harmonic mean. The accuracy, recall, true negative rate and harmonic mean recall for naive bayes and SVM are presented in Table 5.4. The naive bayes classifier has better overall accuracy, but the SVM presents a higher harmonic mean recall.

Chapter 6

Conclusion

The main objective of this project was to develop a conflict strategy prediction system: given a conflict situation description, and a person description, make a prediction of what conflict strategy is the person most likely to choose. My previous worked emphasized the importance of exploring data driven approaches. I cleaned and merged crowd sourced data consisting of responses to hypothetical conflict situations, person descriptions, and labels for the the type of the conflict strategy used. The current version of the data and cleaning, together with instructions on how to run it, is available at:

https://bitbucket.org/pfontain/conflict-data-cleaning

For the predictive system to be viable, the features available should be predictive of conflict strategy and these relations consistent with previous research. The statistical analysis of the data delivered just that, with the following results being statistically significant although with small effect sizes: extroverts have responses that are perceived as active and aggressive; conscientious individuals have responses that are perceived as active; neurotic individuals have responses that are perceived as passive and non aggressive. I raised the data supported hypothesis that physical activity and active responses may be correlated, as well as education level and active label. Lastly, scenario specific features, such as friendship level, appear not to be significantly different. As people obviously react differently in different situations, this could mean that we either have too little context information, or more needs to be extracted. Namely, extracting features from the textual description of the scenario could be an interesting path of research.

Regarding the prediction itself, we trained a naive bayes classifier and support vector machine (SVM). Both classifiers presented higher accuracy than random in the test set. The naives bayes had a higher overall accuracy but at the cost of a lower true negative rate (more active responses than passive). The harmonic mean between recall and true negative rate, weighting both classes equally independent of number os instances, is higher for the SVM. These results give some indication that trying to predict strategy choice tendencies using crowd sourced data could be possible.

Both models could take advantage of textual features of the response, such as word sentiment polarity. The relation between a person's weekly physical activity and type of conflict strategy chosen merits testing in a more targeted study. Finally, conflict resolution is inherently an iterative process in which many strategies may be chosen [Pruitt et al., 1997, p. 162]. Consequently, future work should explore iterative process of strategy choice.

Bibliography

[Abbott et al., 2011] Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., and King, J. (2011). How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, pages 2–11. Association for Computational Linguistics.

[Aylett et al., 2007] Aylett, R., Vala, M., Sequeira, P., and Paiva, A. (2007). Fearnot!–an emergent narrative approach to virtual dramas for anti-bullying education. In *Virtual Storytelling. Using Virtual Reality Technologies for Storytelling*, pages 202–205. Springer.

[Campos et al., 2012] Campos, H., Campos, J., Martinho, C., and Paiva, A. (2012). Virtual agents in conflict. In *Intelligent Virtual Agents*, pages 105–111. Springer.

[Campos et al., 2013] Campos, J., Martinho, C., and Paiva, A. (2013). Conflict inside out: A theoretical approach to conflict from an agent point of view. In *in Proceedings of the Twelfth International Conference on Autonomous Agents and Multiagent Systems*.

[Chang and Lin, 2011] Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

[Cheong et al., 2011] Cheong, Y.-G., Khaled, R., Grappiolo, C., Campos, J., Martinho, C., Ingram, G. P., Paiva, A., and Yannakakis, G. (2011). A computational approach towards conflict resolution for serious games. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 15–22. ACM.

[Costa and McCrae, 1985] Costa, P. and McCrae, R. R. (1985). The neo personality inventory.

[Damasio, 1994] Damasio, A. (1994). Descartes' error: Emotion, reason, and the human brain.

[Dias et al., 2011] Dias, J., Mascarenhas, S., and Paiva, A. (2011). Fatima modular towards an agent architecture with a generic appraisal framework. In *Proceedings of the International Workshop on Standards for Emotion Modeling*.

[Facebook, 2014] Facebook (2014). Facebook reports second quarter 2014 results.

[Field and Hole, 2003] Field, A. P. and Hole, G. (2003). *How to design and report experiments*. Sage Los Angeles, CA.

[Fikes and Nilsson, 1972] Fikes, R. E. and Nilsson, N. J. (1972). Strips: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3):189–208.

[Galgani and Hoffmann, 2011] Galgani, F. and Hoffmann, A. (2011). Lexa: Towards automatic legal citation classification. In *Advances in Artificial Intelligence*, pages 445–454. Springer.

[Gomes and Jhala, 2013] Gomes, P. and Jhala, A. (2013). Ai authoring for virtual characters in conflict. In *Ninth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*.

[Gomes et al., 2013] Gomes, P., Paiva, A., Martinho, C., and Jhala, A. (2013). Metrics for character believability in interactive narrative. In *International Conference on Interactive Digital Storytelling*.

[Gomes et al., 2011a] Gomes, P. F., Martinho, C., and Paiva, A. (2011a). Ive been here before! location and appraisal in memory retrieval. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, pages 1039–1046. IFAAMAS.

[Gomes et al., 2011b] Gomes, P. F., Segura, E. M., Cramer, H., Paiva, T., Paiva, A., and Holmquist, L. E. (2011b). Vipleo and phypleo: Artificial pet with two embodiments. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. ACM.

[Grow et al., 2014] Grow, A., Gaudl, S., Gomes, P., Mateas, M., and Wardrip-Fruin, N. (2014). A methodology for requirements analysis of ai architecture authoring tools. In *To be published in Ninth International Conference on the Foundations of Digital Games (FDG)*.

[Hassan et al., 2010] Hassan, A., Qazvinian, V., and Radev, D. (2010). What's with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1245–1255. Association for Computational Linguistics.

[Hocker and Wilmot, 2013] Hocker, J. L. and Wilmot, W. W. (2013). *Interpersonal Conflict*. WC Brown Company.

[Horn, 1989] Horn, L. (1989). *A natural history of negation*. Chicago University Press.

[Jabbari et al., 2006] Jabbari, S., Allison, B., Guthrie, D., and Guthrie, L. (2006). Towards the orwellian nightmare: separation of business and personal emails. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 407–411. Association for Computational Linguistics.

[Jankowski-Lorek et al., 2014] Jankowski-Lorek, M., Nielek, R., Wierzbicki, A., and Zielinski, K. (2014). Predicting controversy of wikipedia articles using the article feedback tool. In *Proc. Seventh ASE International Conference on Social Computing*.

[Kanno et al., 2006] Kanno, T., Nakata, K., and Furuta, K. (2006). A method for conflict detection based on team intention inference. *Interacting with Computers*, 18(4):747–769.

[Kilmann and Thomas, 1975] Kilmann, R. H. and Thomas, K. W. (1975). Interpersonal conflict-handling behavior as reflections of jungian personality dimensions. *Psychological reports*, 37(3):971–980.

[Mateas, 2002] Mateas, M. (2002). *Interactive drama, art and artificial intelligence*. PhD thesis, Carnegie Mellon University.

[Mateas and Stern, 2004] Mateas, M. and Stern, A. (2004). A behavior language: Joint action and behavioral idioms. *Life-like Characters. Tools, Affective Functions and Applications*, 194:1–28.

[McCoy et al., 2010] McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., and Wardrip-Fruin, N. (2010). Comme il faut 2: a fully realized model for socially-oriented gameplay. In *Proceedings of the Intelligent Narrative Technologies III Workshop*, page 10.

[Misra and Walker, 2013] Misra, A. and Walker, M. A. (2013). Topic independent identification of agreement and disagreement in social media dialogue. In *Proceedings of the SIGDIAL 2013 Conference*, pages 41–50. Association for Computational Linguistics.

[Mitchell, 1997] Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

[Myers, 1962] Myers, I. (1962). *Manual: The Myers-Briggs type indicator*. Educational Testing Service.

[Ortony et al., 1990] Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

[Pruitt et al., 1997] Pruitt, D. G., Parker, J. C., and Mikolic, J. M. (1997). Escalation as a reaction to persistent annoyance. *International Journal of Conflict Management*, 8(3):252–270.

[Pynadath and Tambe, 2002] Pynadath, D. V. and Tambe, M. (2002). Multiagent teamwork: Analyzing the optimality and complexity of key theories and models. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, pages 873–880. ACM.

[Rammstedt and John, 2007] Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1):203–212.

[Shapiro et al., 2013] Shapiro, D., McCoy, J., Grow, A., Samuel, B., Stern, A., Swanson, R., Treanor, M., and Mateas, M. (2013). Creating playable social experiences through whole-body interaction with virtual characters. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.

[Sina et al., 2014] Sina, S., Kraus, S., and Rosenfeld, A. (2014). Using the crowd to generate content for scenario-based serious-games. *arXiv preprint arXiv:1402.5034*.

[Swanson and Jhala, 2012] Swanson, R. and Jhala, A. (2012). A crowd-sourced collection of narratives for studying conflict. *Proceedings of Computational Models of Narrative*, pages 65–73.

[Sycara, 1993] Sycara, K. P. (1993). Machine learning for intelligent support of conflict resolution. *Decision Support Systems*, 10(2):121–136.

[Vasconcelos et al., 2009] Vasconcelos, W. W., Kollingbaum, M. J., and Norman, T. J. (2009). Normative conflict resolution in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 19(2):124–152.

[Walker et al., 2012] Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., and King, J. (2012). A corpus for research on deliberation and debate. In *LREC*, pages 812–817.

[Ware and Young, 2011] Ware, S. G. and Young, R. M. (2011). Cpocl: A narrative planner supporting conflict. In *AIIDE*.